

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Robert Francis Squibbs) RE: Claim to Priority
Serial No.: Not yet assigned)
Filed: Concurrently herewith) Our Ref: B-5257 621329-5
For: "VISUALIZATION OF SPATIALIZED)
AUDIO") Date: October 9, 2003

CLAIM TO PRIORITY UNDER 35 U.S.C. 119

Mail Stop Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450
Sir:

[X] Applicants hereby make a right of priority claim under 35
U.S.C. 119 for the benefit of the filing date(s) of the
following corresponding foreign application(s):

<u>COUNTRY</u>	<u>FILING DATE</u>	<u>SERIAL NUMBER</u>
GB	21 January 2003	0301304.2

[] A certified copy of each of the above-noted patent
applications was filed with the Parent Application No.____.

[X] To support applicants' claim, a certified copy of the above-
identified foreign patent application is enclosed herewith.

[] The priority documents will be forwarded to the Patent Office
when required or prior to issuance.

Respectfully submitted,


Richard P. Berg
Attorney
Reg. No. 28,145

LADAS & PARRY
5670 Wilshire Boulevard
Suite 2100
Los Angeles, CA 90036
Telephone: (323) 934-2300
Telefax: (323) 934-0202



EV30102413-10



INVESTOR IN PEOPLE

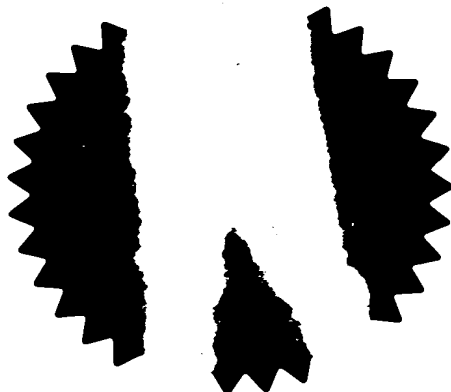
The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



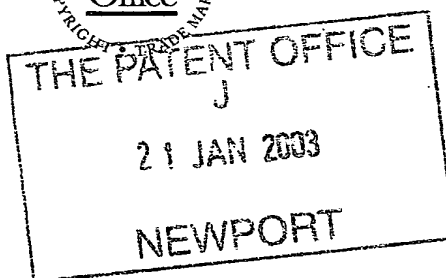
Signed

Dated 29 September 2003



21JAN03 E778544-1 D01463
P01/7700 0.00-0301304.2

EV30102413



The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference 30013357-1 GB

2. Patent application number
(The Patent Office will fill in this part)

0301304.2

3. Full name, address and postcode of the or of each applicant (underline all surnames)

Hewlett-Packard Company
3000 Hanover Street
Palo Alto
CA 94304, USA

Patents ADP number (if you know it)

496588004
Delaware, USA

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention Visualization of Spatialized Audio

5. Name of your agent (if you have one)

"Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)

Robert F. Squibbs
Hewlett-Packard Ltd, IP Section
Filton Road, Stoke Gifford
Bristol BS34 8QZ

Patents ADP number (if you know it)

7928187001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country

Priority application number
(if you know it)

Date of filing
(day / month / year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application

Date of filing
(day / month / year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:

- a) any applicant named in part 3 is not an inventor, or
 - b) there is an inventor who is not named as an applicant, or
 - c) any named applicant is a corporate body.
- See note (d))

Yes

Visualization of Spatialized Audio

Field of the Invention

- 5 The present invention relates to a method and apparatus for providing a visual indication of the likely user-perceived location of one or more sound sources in an audio field generated from left and right audio channel signals.

Background of the Invention

- 10 Methods of acoustically locating a real-world sound source are well known and usually involve the use of an array of microphones; US 5,465,302 and US 6,009,396 both describe sound source location detecting systems of this type. By determining the location of the sound source, it is then possible to adjust the processing parameters of the input from the individual microphones of the array so as to effectively 'focus' the microphone on the
- 15 sound source, enabling the sounds emitted from the source to be picked out from surrounding sounds. However, this prior art is not concerned with the same problem as that addressed by the present invention where the starting point is left and right audio channel signals that have been conditioned to enable the generation of a spatialized sound field to a human user.

20

- It is, of course, well known to process a sound-source signal to form left and right audio channel signals so conditioned that when supplied to a human user via (at least) left and right audio output devices, the sound source is perceived by the user as coming from a particular location; this location can be varied by varying the conditioning of the left and
- 25 right channel signals.

- More particularly, the human auditory system, including related brain functions, is capable of localizing sounds in three dimensions notwithstanding that only two sound inputs are received (left and right ear). Research over the years has shown that localization in
- 30 azimuth, elevation and range is dependent on a number of cues derived from the received sound. The nature of these cues is outlined below.

Azimuth Cues – The main azimuth cues are Interaural Time Difference (ITD – sound on

component, this effect being known as the Franssen effect). Such many-speaker systems are not, however, practical for most situations.

For sound sources that have a fixed presentation (non-interactive), it is possible to produce convincing 3D audio through headphones simply by recording the sounds that would be heard at left and right eardrums were the hearer actually present. Such recordings, known as binaural recordings, have certain disadvantages including the need for headphones, the lack of interactive controllability of the source location, and unreliable elevation effects due to the variation in pinna shapes between different hearers.

10

To enable a sound source to be variably positioned in a 3D audio field, a number of systems have evolved that are based on a transfer function relating source sound pressures to ear drum sound pressures. This transfer function is known as the Head Related Transfer Function (HRTF) and the associated impulse response, as the Head Related Impulse Response (HRIR). If the HRTF is known for the left and right ears, binaural signals can be synthesized from a monaural source. By storing measured HRTF (or HRIR) values for various source locations, the location of a source can be interactively varied simply by choosing and applying the appropriate stored values to the sound source to produce left and right channel outputs. A number of commercial 3D audio systems exist utilizing this principle. Rather than storing values, the HRTF can be modeled but this requires considerably more processing power.

20

The generation of binaural signals as described above is directly applicable to headphone systems. However, the situation is more complex where stereo loudspeakers are used for sound output because sound from both speakers can reach both ears. In one solution, the transfer functions between each speaker and each ear are additionally derived and used to try to cancel out cross-talk from the left speaker to the right ear and from the right speaker to the left ear.

25

Other approaches to those outlined above for the generation of 3D audio fields are also possible as will be appreciated by persons skilled in the art. Regardless of the method of generation of the audio field, most 3D audio systems are, in practice, generally effective in achieving azimuth positioning but less effective for elevation and range. However, in many

30

- (b) detecting corresponding components in the left and right channel signals and using them to infer the presence of at least one sound source and determine its azimuth location; and
- (c) displaying a visual indication of at least one sound source inferred in step (b) and its location.

According to another aspect of the present invention, there is provided apparatus for providing a visual indication of the likely user-perceived location of sound sources in an audio field generated from left and right audio channel signals, the apparatus comprising:

- an input interface for receiving the left and right audio channel signals;
- a correlation arrangement for detecting corresponding components in the left and right channel signals;
- a source-determination arrangement for using the detected corresponding components to infer the presence of at least one sound source and determine its azimuth location;
- and
- a display processing arrangement for causing the display, on a display connected thereto, of a visual indication of at least one sound source inferred in step (b) and its location.

Brief Description of the Drawings

Embodiments of the invention will now be described, by way of non-limiting example, with reference to the accompanying diagrammatic drawings, in which:

- . **Figure 1** is a diagram illustrating the connection of visualization apparatus embodying the invention to a CD player;
- . **Figure 2** is a functional block diagram of the Figure 1 visualization apparatus; and
- . **Figure 3** is a diagram showing the visualization of a focus volume of a 3D audio field experienced by a user having portable audio equipment.

Best Mode of Carrying Out the Invention

Figure 1 shows the connection of visualization apparatus 15 embodying the present invention to a CD player 10. The CD player is a stereo player with left (L) and right (R)

arranged to cause the production of visual indications in respect of all sound sources detected during the course of a sound passage of interest .

Considering the apparatus 15 in more detail, in the present embodiment the input buffers 20 and 21 are digital in form with the left and right audio channel signals received by the apparatus 15 either being digital signals or, if of analogue form, being converted to digital signals by converters (not shown) before being fed to the buffers 20, 21. The buffers 20, 21 are each arranged to hold a half-second segment of the corresponding channel of the sound passage being output by the CD player with the buffers becoming full in correspondence to the end of a processing cycle of the apparatus. At the start of the next processing cycle, the contents of the buffers are transferred to the correlator 22 after which filling of the buffers from the left and right audio channel signals recommences.

The correlator 22 (which is, for example, a digital signal processor) is operative to detect corresponding components by pairing left and right audio-channel tones, potentially offset in time, that match in pitch and in amplitude variation profile. Thus, for example, the correlator 22 can be arranged to sweep through the frequency range of the audio-channel signals and for each tone signal detected in one channel signal, determine if there is a corresponding signal in the other channel signal, potentially offset in time. If a corresponding tone signal is found and it has a similar amplitude variation profile over the time segment being processed, then these left and right channel tone signals are taken as forming a matching pair originating from a common sound source. The matched tones do not, in fact, need to be of a fixed frequency but any frequency variation in one must be matched by the same frequency variation in the other (again, allowing for a possible time offset).

For each matching pair of tones detected by the correlator 22, it feeds an output to a block 24 of the source-determination arrangement 23 giving the characteristic tone frequency (pitch), the average amplitude (across both channels for periods when the tones are present) and the amplitude variation profile of the matched pair; if the pitch of the tone varies, then the initial detected pitch is used for the characteristic pitch. The correlator 22 also outputs to a block 25 of the source-determination arrangement 23, measures of the amplitudes of the matched left and right channel tone signals and/or of their timing offset relative to each

For each group of associated LES records 27 identified by the block 28, a corresponding "located compound sound" (LCS) record 29 is created by block 28 in the memory 26. Each LCS record 29 comprises:

- a LCS ID,
- 5 - an amplitude variation profile formed from a weighted average of the associated LES amplitude variation profiles, the weighting being set to favour the louder LESs (alternatively, for simplification, the amplitude variation profile of the loudest LES can be used instead);
- an harmonic profile giving the relative strengths of the different frequencies of the
- 10 associated LESs as indicated by the average amplitudes recorded in records 27;
- an azimuth location formed from a weighted average of the azimuth locations of the associated LESs, the weighting being set to favour the louder LESs (again, for simplification, the azimuth location of the loudest LES can be taken instead); and
- a last detection timestamp corresponding to the most recent value of the last detection
- 15 timestamps of the associated LESs.

The block 28 may be set to process the LESs created in one operating cycle of the correlator 22 and block 24, in the same operating cycle or in the next following operating cycle; in this latter case, appropriate measures are taken to ensure that block 28 does not try

20 to process LES records being added by block 24 during its current operating cycle.

After the compound-sound identification block 28 has finished determining what LCS are present, a source identification block 30 is triggered to infer and record, for each LCS, a corresponding sound source in a sound source item record 34 stored in a source item

25 memory 33. The block 30 is operative to determine the type of each sound source by matching the harmonic profile and/or amplitude variation profile of the LCS concerned with predetermined sound-source profiles (typically, but not necessarily limited to, musical instrument profiles). Each sound-source item record holds an item ID, the determined sound source type, and the azimuth position and last detection time stamp copied from the

30 corresponding LCS.

Rather than the source identification block 30 carrying out its operation after the block 28 has finished LCS identification, the block can be arranged to create a new sound-source

item and its azimuth location in the audio field. This is preferably done by displaying representations of the sound source items in a spatial relation corresponding to that of the sources themselves. Advantageously, each sound-source representation is indicative of the type of the corresponding sound source, appropriate image data for each type of source item being stored in source item visualization data memory 32 and being retrieved by the display processing stage 35 as needed. The form of representation used can also be varied in dependence on whether the last-detected timestamp recorded for a source item is within a certain time window of the current time; if this is the case then the sound source is assumed to be currently active and a corresponding active image (which may be an animated image) is displayed whereas if the timestamp is older than the window, the sound source is taken to be currently inactive and a corresponding inactive image is displayed.

Rather than all the sound source items being represented at the same time, the display processing stage can be arranged to display only those sound sources that are currently active or that are located within a user-selected portion of the audio field (this portion being changeable by the user). Furthermore, rather than a sound source item having existence from its inception to the end of the sound passage of interest regardless of how long it has been inactive, a sound source item that remains inactive for more than a given period as judged by its last-detected timestamp, can be deleted from the memory 33.

In addition to determining the azimuth location of each detected sound source, the source-determination arrangement 23 can be arranged to determine the depth (radial distance from the user) and/or height location of each sound source. Thus, for example, the depth location of a sound source in the audio field can be determined in dependence on the relative loudness of this sound source as compared to other sound sources. This can conveniently be done by storing in each LCS record 29 the largest average amplitude value of the associated LES records 27, and then arranging for block 30 to use these LCS average amplitude values to allocate depth values to the sound sources.

As regards the height location of a sound source in the audio field, if the audio channel signals have been processed to simulate a pinna notch effect with a view to enabling a human listener to perceive sound source height, then the block 30 can also be arranged to determine the sound source height by assessing the variation with frequency of the relative

this case, the output of the display processing stage 35 would be passed by a wireless link to the display 16.

It will be appreciated that many variants are possible to the above described embodiments of the invention. In particular, the degree of processing effected by the correlator 22 and the source determination arrangement 23 in detecting sound sources can be tailored to the available processing power. For example, rather than every successive audio channel signal segment being processed, only certain segments can be processed, such as every other segment or every third segment. Another processing simplification would be only to consider tones having more than a certain amplitude thereby reducing the processing load concerned with harmonics. Identification of source type can be done simply on the basis of the pitch and amplitude profile and in this case it is possible to omit the identification of "located compound sounds" (LCS) though this is likely to lead to the detection of multiple co-located sources unless provision is made to consolidate such sources into a single source. Determining the type of a sound source item is not, of course, essential. The duration of each audio channel segment can be made greater or less than the half a second described above.

Where ample processing power is available, then the correlator and source determination arrangement can be arranged to operate on a continuous basis rather than on discrete segments.

The above-described functional blocks of the correlator 22 and source-determination arrangement 23 can be implemented in hardware and/or in software. Furthermore, analogue forms of these elements can also be implemented.

inferred as present during any repetition having a continuing existence across at least one subsequent repetition.

6. A method according to claim 4, wherein in the course of a sound passage represented by the left and right audio channel signals, step (b) is carried out repeatedly or on an on-going basis with sound sources inferred as present at any stage having a continuing existence, step (b) involving seeking to match newly-determined compound sounds with known sound sources and only inferring the presence of a new sound source if no such match is possible.

7. A method according to claim 6, wherein in seeking to match newly-determined compound sounds with known sound sources, limited differences in location are allowed between the newly-determined compound sound and a candidate matching sound source the location of which is taken to be that of a previous compound sound associated with the sound source; said limited differences in location serving to allow for movement of the sound source in the audio field.

8. A method according to claim 4, wherein in step (c) at least one sound source inferred as present in step (b) is visually indicated by a visual element representative of the type of sound source.

9. A method according to claim 8, wherein in the course of a sound passage represented by the left and right audio channel signals, step (b) is carried out repeatedly or on an on-going basis with sound sources inferred as present at any stage continuing to be visually represented in step (c) even after the corresponding compound sounds are no longer detected.

10. A method according to claim 9, wherein the visual representation of a said sound source is varied according to whether or not a compound sound corresponding to the sound source has been recently detected.

11. A method according to claim 1, wherein the depth location of a said sound source in the audio field is determined in dependence on the loudness of this sound source, the

16. Apparatus according to claim 15, wherein source-determination arrangement is arranged to associate, into a compound sound, elemental sounds that have the same azimuth location, the same general amplitude variation profile and are harmonically
5 related.

17. Apparatus according to claim 16, wherein the source-determination arrangement is arranged to use the or each compound sound to infer the presence of a corresponding sound source with the type of that sound source being determined according to the harmonic
10 profile and/or amplitude variation profile of the compound sound concerned.

18. Apparatus according to claim 17, wherein the correlation arrangement and source-determination arrangement are arranged such that, in the course of a sound passage represented by the left and right audio channel signals, they carry out their respective
15 functions repeatedly with the elemental and compound sounds being newly determined at each repetition but sound sources inferred as present during any repetition being remembered by the source-determination arrangement across at least one subsequent repetition.

20 19. Apparatus according to claim 17, the correlation arrangement and source-determination arrangement are arranged such that, in the course of a sound passage represented by the left and right audio channel signals, they carry out their respective functions repeatedly or on an on-going basis, the source-determination arrangement being further arranged to remember sound sources inferred as present at any stage and to seek to
25 match newly-determined compound sounds with known sound sources and only infer the presence of a new sound source if no such match is possible.

20. Apparatus according to claim 19, wherein the source-determination arrangement is arranged to permit, in seeking to match newly-determined compound sounds with known
30 sound sources, limited differences in location between the newly-determined compound sound and a candidate matching sound source the location of which is taken to be that of a previous compound sound associated with the sound source.

26. Apparatus according to claim 14, wherein the display processing arrangement is arranged to cause visual indications to be displayed for only those sound sources located within a portion of said audio field, the display processing arrangement including a user-controllable input device for selecting the position of this portion within the audio field.

1 / 3

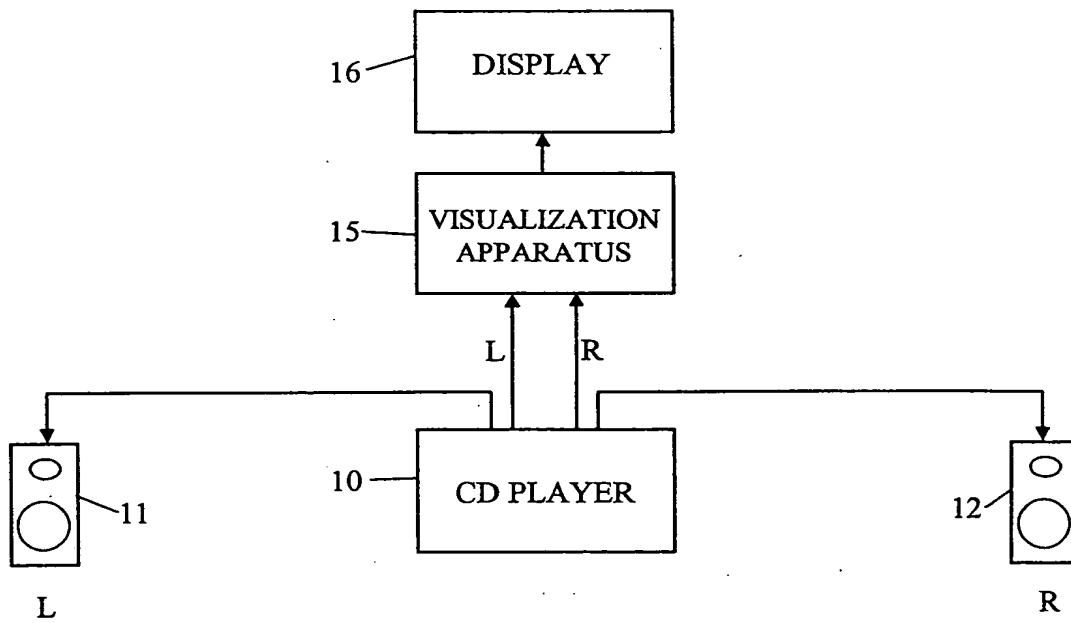


Figure 1

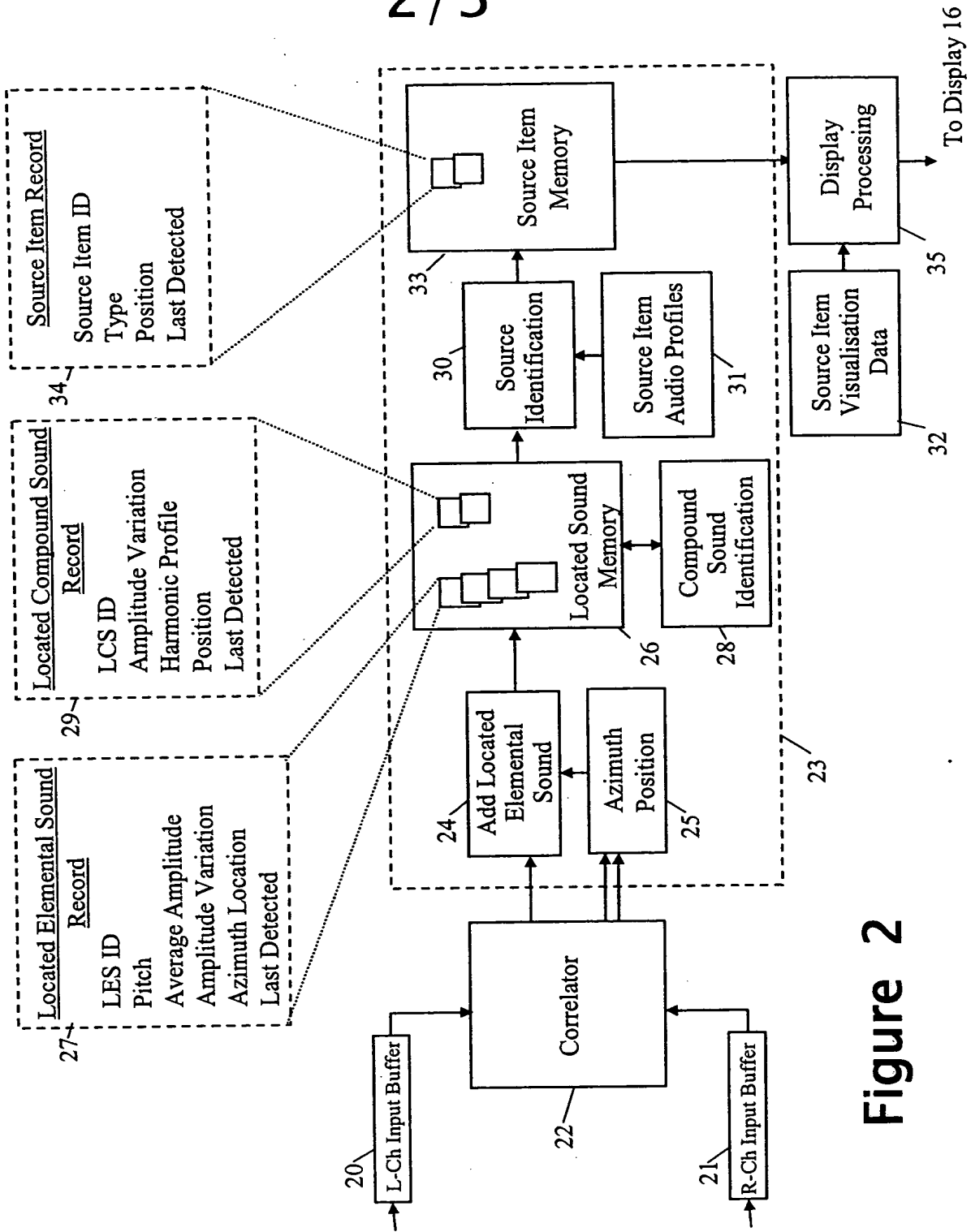


Figure 2

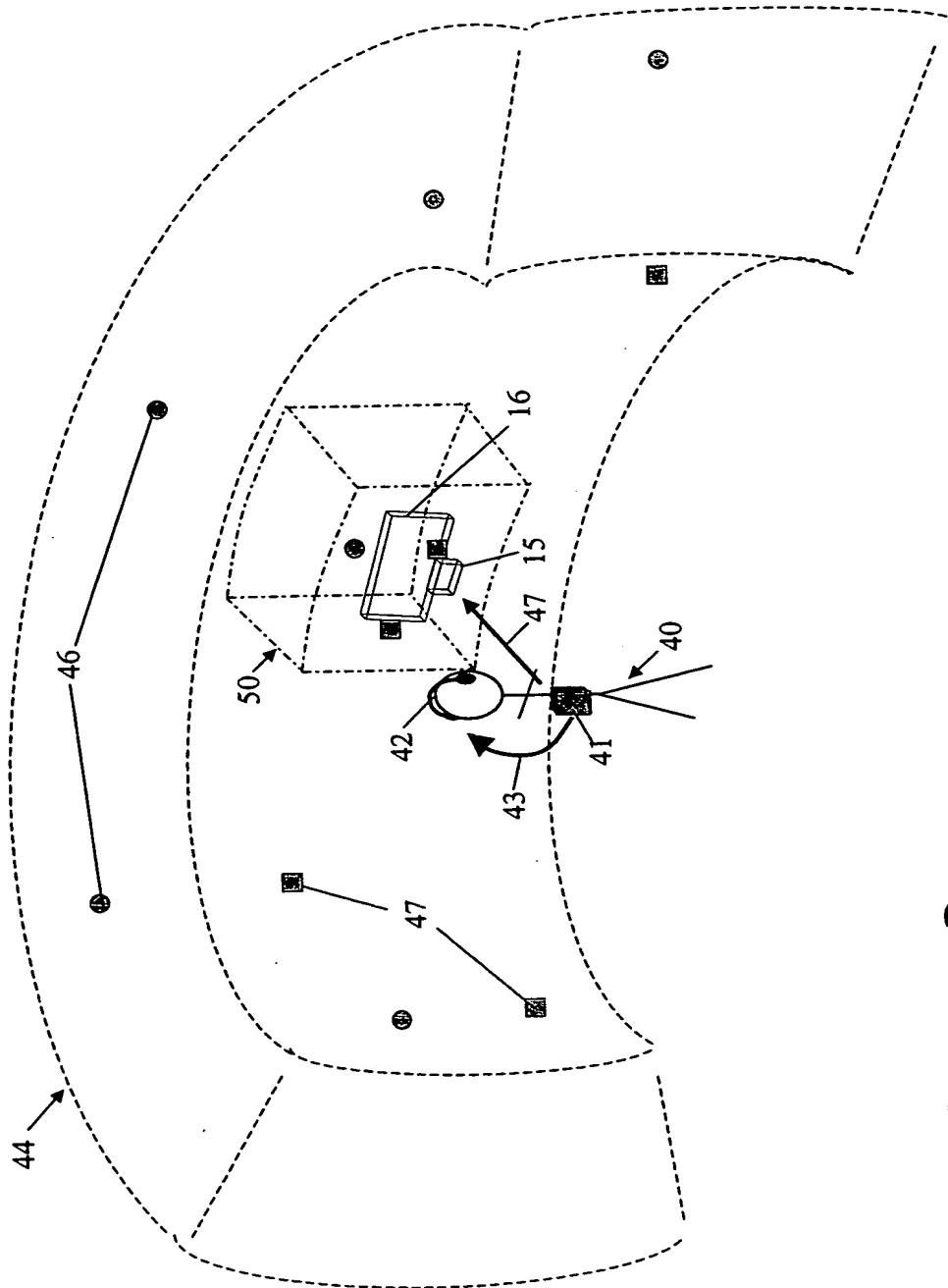


Figure 3